

RECEIVED  
CENTRAL FAX CENTER

MAY 23 2006

**Yee &  
Associates, P.C.**4100 Alpha Road  
Suite 1100  
Dallas, Texas 75244Main No. (972) 385-8777  
Facsimile (972) 385-7766**Facsimile Cover Sheet**

To: Commissioner for Patents for Examiner Laurie Anne Ries Group Art Unit 2176	Facsimile No.: 571/273-8300
From: Carrie Parker Legal Assistant to Mari Stewart	No. of Pages Including Cover Sheet: 31
Message:  Enclosed herewith: <ul style="list-style-type: none"><li>• Transmittal of Appeal Brief; and</li><li>• Appeal Brief.</li></ul>	
Re: Application No. 09/838,428 Attorney Docket No: AUS920010015US1	
Date: Tuesday, May 23, 2006	
Please contact us at (972) 385-8777 if you do not receive all pages indicated above or experience any difficulty in receiving this facsimile.	<i>This Facsimile is intended only for the use of the addressee and, if the addressee is a client or their agent, contains privileged and confidential information. If you are not the intended recipient of this facsimile, you have received this facsimile inadvertently and in error. Any review, dissemination, distribution, or copying is strictly prohibited. If you received this facsimile in error, please notify us by telephone and return the facsimile to us immediately.</i>

**PLEASE CONFIRM RECEIPT OF THIS TRANSMISSION BY  
FAXING A CONFIRMATION TO 972-385-7766.**

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

RECEIVED

CENTRAL FAX CENTER

In re application of: Janakiraman et al.

Serial No.: 09/838,428

Filed: April 19, 2001

For: Displaying Text of Video in  
Browsers on a Frame by Frame basis

Group Art Unit: 2176

Examiner: Ries, Laurie Anne

Attorney Docket No.: AUS920010015US1

MAY 23 2006

35525

PATENT TRADEMARK OFFICE  
CUSTOMER NUMBERCertificate of Transmission Under 37 C.F.R. § 1.8(a)I hereby certify this correspondence is being transmitted via  
facsimile to the Commissioner for Patents, P.O. Box 1450,  
Alexandria, VA 22313-1450, facsimile number (571) 273-8300,  
on May 23, 2006.

By:

Carrie Parker  
Carrie ParkerTRANSMITTAL OF APPEAL BRIEFCommissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

ENCLOSED HERewith:

- Appeal Brief (37 C.F.R. 41.37)

No fee is required for filing an Appeal Brief. Please charge this fee to IBM Corporation Deposit Account No. 09-0447. No additional fees are believed to be necessary. If, however, any additional fees are required, I authorize the Commissioner to charge these fees which may be required to IBM Corporation Deposit Account No. 09-0447. No extension of time is believed to be necessary. If, however, an extension of time is required, the extension is requested, and I authorize the Commissioner to charge any fees for this extension to IBM Corporation Deposit Account No. 09-0447.

Respectfully submitted,

Mari Stewart

Mari Stewart

Registration No. 50,359

Duke W. Yee

Registration No. 34,285

YEE &amp; ASSOCIATES, P.C.

P.O. Box 802333

Dallas, Texas 75380

(972) 385-8777

ATTORNEYS FOR APPLICANTS

RECEIVED  
CENTRAL FAX CENTER

MAY 23 2006

Docket No. AUS920010015US1

PATENT

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: Janakiraman et al.

Serial No. 09/838,428

Filed: April 19, 2001

For: Displaying Text of Video in  
Browsers on a Frame by Frame Basis§  
§  
§  
§  
§  
§  
§

Group Art Unit: 2176

Examiner: Ries, Laurie Anne

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

35525

PATENT TRADEMARK OFFICE  
CUSTOMER NUMBERCertificate of Transmittal Under 37 C.F.R. § 1.8(a)

I hereby certify this correspondence is being transmitted via  
facsimile to the Commissioner for Patents, P.O. Box 1450,  
Alexandria, VA 22313-1450, facsimile number (571) 273-8300  
on May 23, 2006.

By:

*Carrie Parker*  
Carrie Parker

APPEAL BRIEF (37 C.F.R. 41.37)

This brief is in furtherance of the Notice of Appeal, filed in this case on March 27, 2006.

No fee is required for filing an Appeal Brief. Please charge this fee to IBM Corporation Deposit Account No. 09-0447. No additional fees are believed to be necessary. If, however, any additional fees are required, I authorize the Commissioner to charge these fees which may be required to IBM Corporation Deposit Account No. 09-0447. No extension of time is believed to be necessary. If, however, an extension of time is required, the extension is requested, and I authorize the Commissioner to charge any fees for this extension to IBM Corporation Deposit Account No. 09-0447.

(Appeal Brief Page 1 of 29)  
Janakiraman et al. - 09/838,428

**REAL PARTY IN INTEREST**

The real party in interest in this appeal is the following party: International Business Machines Corporation.

**RELATED APPEALS AND INTERFERENCES**

With respect to other appeals or interferences that will directly affect, or be directly affected by, or have a bearing on the Board's decision in the pending appeal, there are no such appeals or interferences.

**STATUS OF CLAIMS****A. TOTAL NUMBER OF CLAIMS IN APPLICATION**

Claims in the application are: 1, 3-8, 10-15, and 17-21

**B. STATUS OF ALL THE CLAIMS IN APPLICATION**

1. Claims canceled: 2, 9, and 16
2. Claims withdrawn from consideration but not canceled: NONE
3. Claims pending: 1, 3-8, 10-15, and 17-21
4. Claims allowed: NONE
5. Claims rejected: 1, 3-8, 10-15, and 17-21
6. Claims objected to: NONE

**C. CLAIMS ON APPEAL**

The claims on appeal are: 1, 3-8, 10-15, and 17-21.

**STATUS OF AMENDMENTS**

There are no amendments after final rejection. Therefore, claims 1, 3-8, 10-15, and 17-21 are as amended in the last submitted Response to Office Action filed on September 29, 2005.

**SUMMARY OF CLAIMED SUBJECT MATTER*****Independent claims 1, 8, and 15:***

The presently claimed invention provides a method, computer program product, and system for presenting text from moving video to a user. The present invention receives multimedia data containing a plurality of moving video frames and an associated plurality of sets of text data (see specification at page 13, lines 23-29; page 14, lines 16-21; page 16, lines 30-32; and page 19, lines 5-16), wherein the associated plurality of sets of text data are associated in time with the plurality of moving video frames (see specification at page 15, lines 2-8 and Figure 7), wherein the plurality of sets of text data includes a first text data set associated with a first plurality of moving video frames and a second text data set associated with a second plurality of moving video frames (see specification, page 11, line 24, to page 12, line 15; page 16, line 26, to page 17, line 2). The present invention extracts the associated plurality of sets of text data from the multimedia data (see specification, page 11, lines 15-23; page 13, lines 23-31; page 14, line 25, to page 16, line 25; page 17, lines 3-12). The present invention extracts a first video frame from the first plurality of moving video frames associated with the first text data set to form a first still image (see specification at page 11, lines 24-32; page 12, lines 12-26). The present invention extracts a second video frame from the second plurality of moving video frames associated with the first text data set to form a second still image (see specification at page 11, lines 24-32; page 12, lines 12-26; page 15, lines 2-8; page 17, lines 18-22 and Figure 5). The present invention outputs the first text data set in association with the first still image (see specification, page 20 lines 19-24). The present invention outputs the second text data set in association with the second still image (see specification, page 11, line 24, to page 12, line 15; page 17, lines 12-22; page 20, lines 19-24; and Figure 8).

The means recited in independent claim 15, as well as dependent claims 17-21, may be data processing hardware within server 200, client 300, and combinations thereof, as described in the specification at page 6, line 2, to page 10, line 20, operating under control of software performing with the functionality described in the specification at page 10, line 21, to page 14, line 12, or equivalent.



**GROUND OF REJECTION TO BE REVIEWED ON APPEAL**

The grounds of rejection on appeal are as follows:

**A. GROUND OF REJECTION 1 (Claims 1, 3-6, 8, 10-13, 15, and 17-20)**

The grounds of rejection on appeal are as follows:

Claims 1, 3-6, 8, 10-13, 15, and 17-20 are rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over *Loui*. (U.S. Patent No. 6, 813, 618 B1) in view of *Bergen*. (U.S. Patent No. 6, 956, 573 B1).

**B. GROUND OF REJECTION 1 (Claims 7, 14, and 21)**

Claims 7, 14, and 21 are rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over *Loui*. (U.S. Patent No. 6, 813, 618 B1) in view of *Bergen*. (U.S. Patent No. 6, 956, 573 B1) and further in view of *Cruz* ("A User-Centered Interface for Querying Distributed Multimedia Databases").

### **ARGUMENT**

#### **A. 35 U.S.C. § 103, Alleged Obviousness, Claims 1, 3-6, 8, 10-13, 15, and 17-20**

The Final Office Action rejects claims 1, 3-6, 8, 10-13, 15, and 17-20 under 35 U.S.C. § 103(a) as being allegedly unpatentable over *Loui*. (U.S. Patent No. 6, 813, 618 B1) in view of *Bergen*. (U.S. Patent No. 6, 956, 573 B1). This rejection is respectfully traversed.

##### **1. The examiner bears the burden of establishing a *prima facie* case of obviousness.**

The Examiner bears the burden of establishing a *prima facie* case of obviousness based on the prior art when rejecting claims under 35 U.S.C. § 103. In *re Fritch*, 972 F.2d 1260, 23 U.S.P.Q.2d 1780 (Fed. Cir. 1992). In this case, the examiner has failed to establish a *prima facie* case of obviousness because the cited references do not teach the features of the present invention as believed by the examiner and the references cannot be properly modified or combined to reach the presently claimed invention for the reasons stated below.

*Loui* teaches a system for acquisition of related graphical material in a digital graphics album. *Loui* adds graphical material, such as digital images, to a digital graphics album. *Loui* states:

Reference material in a digital graphics album is specified. Annotation data is extracted from the reference material and may be processed by a natural language processor to produce search keywords. In addition to the keywords, user directives may be provided, both of which are used to conduct a search for related graphical materials. The search is conducted by querying a graphical material database through a network connection. The search results are received and the user can select from the resultant materials for inclusion in the digital graphics album. If no satisfactory material is found, the user can specify a reference graphical image that is processed to produce search criteria that are image content descriptors. The database is again queried in accordance with these descriptors to provide search results for possible inclusion.

*Loui*, Abstract.

*Loui* teaches searching for graphical images based on a keyword search or image content descriptors. If any related graphical materials are found, the resultant materials can be selected

by a user for inclusion in the user's graphical image album. Thus, *Loui* merely teaches a system for adding graphical images to a graphical images album.

*Bergen* is directed to a system that facilitates efficiently representing, storing, and accessing video information. *Bergen* teaches:

A method and concomitant apparatus for comprehensively representing video information in a manner facilitating indexing of the video information. Specifically, a method according to the invention comprises the steps of dividing a continuous video stream into a plurality of video scenes; and at least one of the steps of dividing, using intra-scene motion analysis, at least one of the plurality of scenes into one or more layers; representing, as a mosaic, at least one of the plurality of scenes; computing, for at least one layer or scene, one or more content-related appearance attributes; and storing, in a database, the content-related appearance attributes or said mosaic representations.

*Bergen*, Abstract.

As shown above, *Bergen* segments video information into scenes. The video may be divided into scenes based on intra-scene motion analysis. Thus, *Bergen* merely describes representing video information in a manner that facilitates indexing of the video information.

In contradistinction, the presently claimed invention in claim 1 is concerned with providing a method, computer program product, and system for presenting text associated with moving video. The present invention extracts a plurality of sets of text data from multimedia data containing a plurality of moving video frames, extracts video frames associated with the sets of text data to form still images, and outputs the sets of text data in association with the still images.

All claim limitations must be considered, especially when missing from the prior art. In comparing *Loui* and *Bergen* to the claimed invention, the claim limitations of the presently claimed invention may not be ignored in an obviousness determination. Independent claim 1 recites as follows:

1. A method for presenting text from moving video to a user, the method comprising:  
receiving multimedia data containing a plurality of moving video frames and an associated plurality of sets of text data, wherein the associated plurality of sets of text data are associated in time with the plurality of moving video frames, wherein the plurality of sets of text data includes a first text data set associated with a first plurality of moving video frames of the multimedia data, and a second

text data set associated with a second plurality of moving video frames of the multimedia data;  
extracting the associated plurality of sets of text data from the multimedia data;  
extracting a first video frame, from the first plurality of moving video frames, associated with the first text data set to form a first still image;  
extracting a second video frame, from the second plurality of moving video frames, associated with the first text data set to form a second still image;  
outputting the first text data set in association with the first still image;  
and  
outputting the second text data set in association with the second still image.

Independent claims 8 and 15 recite similar subject matter.

*Loui* and *Bergen*, taken either alone or in combination, fails to teach or suggest the feature of a plurality of moving video frames and an associated plurality of sets of text data, wherein the associated plurality of sets of text data are associated in time with the plurality of moving video frames, wherein the plurality of sets of text data includes a first text data set associated with a first plurality of moving video frames of the multimedia data, and a second text data set associated with a second plurality of moving video frames of the multimedia data, as is recited in claim 1.

In addition, *Loui* and *Bergen*, taken either alone or in combination fails to teach or suggest the steps for extracting the associated plurality of sets of text data from the multimedia data; extracting a first video frame, from the first plurality of moving video frames, associated with the first text data set to form a first still image; and extracting a second video frame, from the second plurality of moving video frames, associated with the first text data set to form a second still image, as is also claimed in independent claim 1.

**Loui**

The Examiner acknowledges that *Loui* does not disclose that the video frames or still images are captured from moving video. Because *Loui* does not teach moving video, *Loui* cannot possibly teach or suggest "a plurality of moving video frames and an associated plurality of sets of text data, wherein the associated plurality of sets of text data are associated in time with the plurality of moving video frames, wherein the plurality of sets of text data includes a first text data set associated with a first plurality of moving video frames of the

multimedia data, and a second text data set associated with a second plurality of moving video frames of the multimedia data, as is recited in claim 1. For example, the Examiner alleges that *Loui* discloses the associated number of sets of text data are associated in time with the number of video frames at column 2, lines 1-5, which states as follows:

Modern camera systems have evolved and some now provide a means of generating annotation data for digital graphic images. Cameras may have a built in clock which time stamps the images. Some allow entry of textual data that can be associated with the digital images. Some even include a global position systems (GPS) receiver which can mark images with the geographic location of the camera at the time the image is exposed. Some allow for voice annotation. All of these kinds of information can be fed to the digital graphics alburning application to be used to annotate the digital graphics materials.

*Loui*, column 2, lines 1-11.

Here, *Loui* describes cameras having a built-in clock to time stamp an image. *Loui* merely describes various kinds of information fed to a digital graphics alburning application to annotate digital graphic images inserted into a graphics album, such as a time or location of a camera when the image is exposed. However, a time stamp on a digital image records a time that a given image was taken. A time stamp does not teach or suggest sets of text data having a time association with moving video frames, as is claimed in claim 1. Thus, *Loui* does not teach or suggest "a plurality of moving video frames and an associated plurality of sets of text data, wherein the associated plurality of sets of text data are associated in time with the plurality of moving video frames, wherein the plurality of sets of text data includes a first text data set associated with a first plurality of moving video frames of the multimedia data, and a second text data set associated with a second plurality of moving video frames of the multimedia data," in this or any other section of the reference.

Moreover, because *Loui* does not teach sets of text data associated in time with the plurality of moving video frames, *Loui* cannot teach or suggest "extracting a first video frame, from the first plurality of moving video frames, associated with the first text data set to form a first still image" and "extracting a second video frame, from the second plurality of moving video frames, associated with the first text data set to form a second still image," as is also claimed in claim 1. The Examiner states that extracting a first video frame, from the number of video frames, associated with the first text data set to form a first still image is disclosed by *Loui*

at column 1, lines 61-65 and column 6, lines 33-37. The cited portion of *Loui* at column 1, lines 61-65 is included in the following paragraph of *Loui* which states:

As a user builds a digital graphic album, there are many choices as to how the images will be organized and annotated. Naturally, digital graphic album software applications allow the user to do this manually. But because of the power of computers and software, software suppliers have added features which make organization of images in digital graphic albums more automated, easier and more flexible. In addition, the kinds of things that can be stored in a digital graphics album has increased. For example, video clips can be placed in the album as well as still images, computer generated graphics, and other digital materials. In the case of a video image, typically a key frame is selected for static display, identifying the video. When a user desires to watch the video, the key frame is selected and this causes the software application to play the video clip.

*Loui*, column 1, lines 61-65.

This portion of *Loui* describes a digital graphic album that can store video clips as well as still images. A key frame is selected for display in the digital graphic album. When a user wants to watch the video stored in the album, the user selects the key frame. Thus, *Loui* merely teaches displaying a selected key frame or still image from a video clip for display in a digital graphic album rather than extracting the still image or key frame from the video clip. In contradistinction, the presently claimed invention in claim 1 extracts a first video frame from the first plurality of moving video frames associated with the first text data set to form a first still image.

The Examiner also cites to *Loui* at column 5, lines 41-49, which is included in the portion of *Loui* that states as follows:

Reference is directed to FIG. 3 which is a diagram of the display in which a user specifies reference material in the digital graphics album. The display 20 appears on the screen of a personal computer. The display 20 has a pull-down menu 24 in this illustrative embodiment. The albuming application has multiple album pages 22 that appear on the screen 20. On the front page, in this example, four graphic materials appear 28 and 26, each of which as some annotation 27 associated therewith. In one illustrative embodiment, if the graphic materials are digital photographs, and the annotation is a brief description of the event in the photograph.

*Loui*, column 5, lines 37-49.

*Loui* describes graphic materials in a digital graphics album having some annotation associated with the material. For example, if the graphic material is a photograph, the annotation is a brief

description of the event in the photograph. *Loui* teaches a digital photograph displayed in a digital graphics album having annotations rather than extracting a first video frame from a first plurality of moving video frames associated with the first text data set to form a first still image.

The Examiner also alleges that *Loui* discloses extracting a second video frame from the second number of video frames associated with the first text data set to form a second still image at column 1, lines 61-65 and column 6, lines 33-37. As discussed above, *Loui* at column 1, lines 61-65, which is shown above, merely describes storing a video clip in a digital graphics album. This section of *Loui* teaches displaying a key frame in a digital graphics album rather than extracting a video frame from a plurality of moving video frames associated with the first text data set. The other cited section of *Loui* at column 6, lines 33-37 is included in the portion of *Loui* that states:

Considering again the range of options 36 offered to the user, in this example the options are: MORE IMAGES LIKE THESE which will cause the processor to prioritize and augment the search to produce results similar to the annotation keywords; IMAGES WITH MORE DETAILS which will cause the processor to prioritize and augment the keywords to produce search results producing detailed images similar to those references selected; IMAGES WITH WIDER VIEWS which will cause the processor to prioritize and augment the keywords to produce resultant images with more expansive views; and IMAGES THAT CONTRAST which will cause the processor to prioritize and augment the keywords to produce search results that are in contrast with the selected reference materials.

*Loui*, column 6, lines 33-46.

This section of *Loui* describes a range of options to prioritize and augment a search for images. Among the options described is a "More Images Like These" option to search for results similar to the annotation keyword. Although *Loui* describes searching for more images similar to the annotation keyword, such a keyword search cannot teach or suggest extracting a second video frame from a second plurality of moving video frames associated with the first text data set to form a second still image, as is claimed in claim 1.

Furthermore, *Loui* does not teach or suggest extracting the associated plurality of sets of text data from the multimedia data. The Examiner alleges this feature is disclosed by *Loui* at column 5, lines 41-49, which is quoted above. As shown above, this section of *Loui* describes graphic material in a digital graphics album having some annotations associated with the graphic

materials. Although *Loui* may describe graphic material having associated annotations, such descriptions do not teach or suggest extracting the associated plurality of sets of text data from the multimedia data, where the associated plurality of sets of text data are associated in time with the plurality of moving video frames contained in the multimedia data, as is claimed in claim 1.

The Examiner also cites to *Loui* at column 6, lines 33-37, which is quoted above. As discussed above, this section of *Loui* merely describes an option for a user to prioritize and augment a search for images to produce results similar to an annotation keyword. The keyword search described by *Loui* cannot expressly or impliedly teach or suggest extracting an associated plurality of sets of text data from multimedia data that contains a plurality of moving video frames and the associated plurality of sets of text data.

Moreover, as discussed above, *Loui* does not teach or suggest that video frames or still images are extracted from moving video frames, as is also claimed in claim 1. As shown above, *Loui* merely teaches is an album where keywords are associated with graphic images and searching for additional images using keywords. *Loui* does not teach or suggest extracting the associated plurality of sets of text data from the multimedia data, extracting a first video frame from a first plurality of moving video frames associated with the first text data set, or extracting a second video frame from the second plurality of moving video frames associated with the first text data set in this or any other section of the reference.

### **Bergen**

*Bergen* fails to make up for the deficiencies of *Loui*. The Examiner alleges *Bergen* discloses dividing a continuous video stream into a number of scenes in the Abstract, which is shown above. As discussed above, the cited portion of *Bergen* teaches dividing a continuous video stream based on intra-scene motion analysis. *Bergen* does not teach or suggest dividing a video stream based on sets of text data associated in time with moving video frames. Therefore, *Bergen* does not make up for the deficiencies of *Loui*.

## **2. A proper prima facie case of obviousness must be supported by some teaching or suggestion contained in the prior art.**

A proper prima facie case of obviousness must be supported by some teaching or suggestion contained in the combined references. Applicant respectfully submits that the references cited cannot be combined to produce the claimed invention. The rule is: Obviousness



cannot be established by combining the teachings of the prior art to produce the claimed invention absent some teaching, suggestion or incentive supporting the combination. In re Geiger, 815 F.2d 686, 688, 2 U.S.P.Q.2d 1276, 1278 (Fed. Cir. 1987)(emphasis added).

*Loui* does not give any teaching, suggestion, or incentive to extract a plurality of sets of text data associated in time with a plurality of moving video frames from multimedia data. *Loui* teaches an album where keywords are associated with graphic images. *Loui* does not actually extract sets of text data that are associated in time with any moving video frames. Furthermore, *Loui* does not provide any teaching, suggestion or incentive to extract a first or second video frame from the plurality of moving video frames associated with the first text data set to form a still image, as in the presently claimed invention. *Loui* only teaches storing a video clip in an album and using a key frame for static display in the album to select the video clip when a user wants to watch the video. No suggestion of a combination of components necessary to extract sets of text data associated in time with moving video frames is found in *Loui*. Furthermore, the Examiner has not pointed out any teaching, suggestion, or incentive provided by *Loui* to extract sets of text data associated in time with moving video frames.

Furthermore, *Bergen* does not provide any teaching, suggestion, or incentive to sets of text data associated in time with moving video frames, as in the presently claimed invention. As shown above, *Bergen* is directed towards efficiently representing, storing, and accessing video information. *Bergen* teaches dividing a continuous video stream based on intra-scene motion analysis. Extracting sets of text data associated with the video stream would serve no useful purpose in indexing the video information either before or after dividing the video stream. Thus, *Bergen* does not provide any teaching, suggestion, or motivation to extract sets of text data associated in time with moving video frames or extract a video frame from the moving video frames associated with the first text data set to form a still image. The Examiner has not pointed out any teaching, suggestion, or incentive in *Bergen* to combine or modify *Bergen* to extract sets of text data associated in time with moving video frames or extract a video frame from the moving video frames associated with the first text data set to form a still image.

3. Stating that it is obvious to try or make a modification or combination without a suggestion in the prior art is not *prima facie* obviousness.

The mere fact that a prior art reference can be readily modified does not make the modification obvious unless the prior art suggested the desirability of the modification. In re Laskowski, 871 F.2d 115, 10 U.S.P.Q.2d 1397 (Fed. Cir. 1989) and also see In re Fritch, 972 F.2d 1260, 23 U.S.P.Q.2d 1780 (Fed. Cir. 1992) and In re Mills, 916 F.2d 680, 16 U.S.P.Q.2d 1430 (Fed. Cir. 1993). The examiner may not merely state that the modification would have been obvious to one of ordinary skill in the art without pointing out in the prior art a suggestion of the desirability of the proposed modification. The Examiner states that it would have been obvious to a person of ordinary skill in the art to extract the video frames or still images of *Loui* from the continuous video stream of *Bergen*. The Examiner alleges the motivation for doing so would have been to provide scene-based information from the video to a user. The Examiner cites to *Bergen* at column 2, lines 29-32 which states:

The invention is directed toward providing an information database suitable for providing a scene-based video information to a user. The representation may include motion or may be motionless, depending on the application.

*Bergen*, column 2, lines 29-32.

Here, *Bergen* states that the invention is directed toward providing an information database for providing scene-based video information to a user. As discussed earlier, *Bergen* accomplishes this by dividing a continuous video stream into a plurality of video scenes. The Examiner believes it would have been obvious to combine *Bergen* with *Loui* for the benefit of providing scene-based information from the video to a user to obtain the invention as specified in claims 1, 8, and 15. However, the cited portion of *Bergen* does not suggest that the reference should be modified or combined in the manner suggested by the Examiner. Moreover, even if the reference did provide a motivation to provide scene-based information from the continuous video stream to a user, such a benefit would not motivate one of ordinary skill in the art to modify *Loui* and *Bergen* to extract sets of text data associated in time with the moving video frames in the video stream; extract a video frame from the moving video frames associated with the first text data set to form a first still image; and extract a second video frame from the moving video frames associated with the first text data set to form a second still image, as specified in claim 1. Therefore, the Examiner has failed to point out any teaching, suggestion, or motivation to combine and/or modify *Loui* and *Bergen* in the manner necessary to reach the presently claimed invention in claim 1.

4. The proposed modification of the references would not be made when each the references are considered as a whole.

"It is impermissible within the framework of section 103 to pick and choose from any one reference only so much of it as will support a given position, to the exclusion of other parts necessary to the full appreciation of what such reference fairly suggests to one of ordinary skill in the art." In re Hedges, 228 U.S.P.Q. 685, 687 (Fed. Cir. 1986). The present invention in claim 1 is directed towards solving the problem of presenting moving video with associated text. When text is associated in time with moving video, certain users may have difficulties reading the text within the time constraints of the video. Also, for some users, the moving video may be distracting. Thus, the presently claimed invention extracts sets of text data associated in time with moving video frames from multimedia data. The presently claimed invention in claim 1 outputs the extracted sets of text data in association with still images, rather than moving video.

Neither *Loui* nor *Bergen* teaches or suggests extracting an associated plurality of sets of text data from multimedia data, extracting video frames from the first plurality of moving video frames to form still images, and outputting the sets of text data in association with the still images. In fact, *Loui* and *Bergen* do not even recognize the problem or its source. *Loui* is directed toward solving the problem of searching and selecting digital images for use in digital graphics albums. *Loui* teaches:

An aspect of the subsequent arrangements that a user may make to a photo album is that the user may desire to add additional images to complete the album. As was discussed earlier, the sources are many and varied. This presents a problem to the user because the user may know what kind of images are desired, but not know where to obtain such images. For example, suppose a user has returned from a vacation in France and has a collection of images and videos from the vacation. These are placed in the digital graphics album, annotated and arranged. Upon review, the user realizes that there are several images of the user in the vicinity of the Eiffel Tower, but no images of the Tower itself. Or perhaps the user knows that during the vacation, a major news story broke about France, and the users desires a video clip for the album. Through some amount of search, the user may find such digital graphics materials, but such searching is cumbersome and time consuming.

Consequently, a need exists in the art for an automatic way of identifying, searching and selecting digital graphical materials for use in supplementing digital graphics albums.

*Loui*, column 2, lines 31-50.

Thus, *Loui* is concerned with searching for graphical images for use in an album. *Loui* solves this problem by performing a keyword search and/or searching using image content descriptors to locate desired images. *Loui* states:

The need in the art is addressed by the apparatus and methods of the present invention. In an illustrative embodiment of the present invention, a method of adding graphical material to a digital graphics album is disclosed. The method includes specifying reference material in a digital graphics album and extracting annotation data from said reference material. Then, processing the extracted annotation data by a natural language processor to produce search keywords. User directive data is then received and processed by the natural language processor to produce additional keywords. Both the keywords and additional keywords are prioritized followed by querying a graphical material database through a network connection in accordance with the keywords. Then, receiving from the database at least one resultant graphical material and selecting one or more of the resultant graphical material for insertion into the digital graphics album. However, if none of the resultant graphical materials is selected, specifying at least one reference graphic material indicative of a desired search result and processing the reference graphical material to produce search criteria that are image content descriptors. Using the image content descriptors, querying an image content database through a network connection, and receiving from the image content database at least one resultant image. Having received the resultant image or images, selecting at least one of the resultant images, and inserting the selected resultant image in the digital graphics album.

*Loui*, column 2, line 54-column 3, line 13.

Thus, *Loui* solves the problem of searching for digital images for an album by producing search keywords and/or image content descriptors to search an image content database for graphical images to insert into a graphics album. *Loui* provides a complete solution to the problem. *Loui* does not provide any teaching, suggestion, or motivation to combine or modify the reference to extract sets of text data from multimedia data, extract video frames from the first plurality of moving video frames to form still images, and outputting the sets of text data in association with the still images.

Moreover, *Bergen* is directed towards solving the problems associated with representing, storing, and accessing video information. *Bergen* states:

The capturing of analog video signals in the consumer, industrial and government/military environments is well known. For example, a moderately priced personal computer including a video capture board is typically capable of converting an analog video input signal into a digital video signal, and storing the digital video signal in a mass storage device (e.g., a hard disk drive). However,

the usefulness of the stored digital video signal is limited due to the sequential nature of present video access techniques. These techniques treat the stored video information as merely a digital representation of a sequential analog information stream. That is, stored video is accessed in a linear manner using familiar VCR-like commands, such as the PLAY, STOP, FAST FORWARD, REWIND and the like. Moreover, a lack of annotation and manipulation tools due to, e.g., the enormous amount of data inherent in a video signal, precludes the use of rapid access and manipulation techniques common in database management applications.

Therefore, a need exists in the art for a method and apparatus for analyzing and annotating raw video information to produce a video information database having properties that facilitate a plurality of non-linear access techniques.

*Bergen*, column 1, lines 14-37.

*Bergen* solves the need for analyzing and annotating video information by dividing continuous video stream into a plurality of video scenes using intra-scene motion analysis, representing at least one of the scenes as a mosaic, computing content-related appearance attributes, and storing the content-related appearance attributes or mosaic representations in a database. *Bergen* states:

The invention is a method and apparatus for comprehensively representing video information in a manner facilitating indexing of the video information. Specifically, a method according to the invention comprises the steps of dividing a continuous video stream into a plurality of video scenes; and at least one of the steps of dividing, using intra-scene motion analysis, at least one of the plurality of scenes into one or more layers; representing, as a mosaic, at least one of the plurality of scenes; computing, for at least one layer or scene, one or more content-related appearance attributes; and storing, in a database, the content-related appearance attributes or said mosaic representations.

*Bergen*, column 1, lines 41-52.

Thus, *Bergen* provides a complete solution to the problem of representing, storing, and accessing video information. *Bergen* does not provide any teaching, suggestion, or motivation to modify or combine *Bergen* in the manner necessary to reach the presently claimed invention in claim 1 when *Bergen* is considered as a whole. Therefore, one of ordinary skill in the art would not be motivated to make the examiner's proposed combination and modifications to reach the presently claimed invention when *Loui* and *Bergen* are considered as a whole.

Moreover, the examiner may not use the claimed invention as an "instruction manual" or "template" to piece together the teachings of the prior art so that the invention is rendered

obvious. In re Fritch, 972 F.2d 1260, 23 U.S.P.Q.2d 1780 (Fed. Cir. 1992). Such reliance is an impermissible use of hindsight with the benefit of applicant's disclosure. *Id.* Therefore, absent some teaching, suggestion, or incentive in the prior art, *Loui* and *Bergen* cannot be properly combined to form the claimed invention. As a result, absent any teaching, suggestion, or incentive from the prior art to make the proposed combination, the presently claimed invention can be reached only through the impermissible use of hindsight with the benefit of applicant's disclosure a model for the needed changes.

Thus, *Loui* and *Bergen*, taken alone or in combination, fail to teach or suggest all of the features in independent claim 1. Independent claim 8 and 15 recite subject matter addressed above with respect to claim 1 and are allowable for similar reasons. At least by virtue of their dependency on claims 1, 8, and 15, the specific features of claims 3-6, 10-13, and 17-20 are not taught or suggested by *Loui* and *Bergen*, wither alone or in combination. Accordingly, Appellants respectfully request that the rejection of claims 1, 3-6, 8, 10-13, 15, and 17-20 under 35 U.S.C. § 103(a) not be sustained.

**B. 35 U.S.C. § 103, Alleged Obviousness, Claims 7, 14, and 21**

The Final Office Action rejects claims 7, 14, and 21 are rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over *Loui*. (U.S. Patent No. 6, 813, 618 B1) in view of *Bergen*. (U.S. Patent No. 6, 956, 573 B1) and further in view of *Cruz* ("A User-Centered Interface for Querying Distributed Multimedia Databases"). The rejection is respectfully traversed.

Claims 7, 14, and 21 are dependent on independent claims 1, 8, and 15. Thus, these claims are not obvious over *Loui* in view of *Bergen* for at least the reasons noted above with regards to claims 1, 8, and 15. Moreover, *Cruz* does not provide for the deficiencies of *Loui* and *Bergen* and, thus, any alleged combination of *Loui*, *Bergen*, and *Cruz* would not be sufficient to reject independent claims 1, 8, and 15 or claims 7, 14, and 21 by virtue of their dependency. That is, *Cruz* does not teach or suggest the a plurality of moving video frames and an associated plurality of sets of text data associated in time with the plurality of moving video frames; extracting sets of text data from the multimedia data; and extracting a video frame from the first plurality of moving video frames associated with the first text data set to form a still image.

*Cruz* is directed toward the problem of finding relevant information in the vastly growing realm of digital media. *Cruz* states:

Facilitating information retrieval in the vastly growing realm of digital media has become increasingly difficult. Delaunay<sup>MM</sup> seeks to assist all users in finding relevant information through an interactive interface that supports pre- and post-query refinement, and a customizable multimedia information display. This project leverages the strengths of visual query languages with a resourceful framework to provide users with a single intuitive interface. The interface and its supporting framework are described in this paper.

*Cruz*, Abstract.

As shown above, *Cruz* solves the problem of querying multimedia databases. *Cruz* is unconcerned with the problems associated with moving video with associated text where certain users have difficulty reading the text within the time constraints of the video and where the moving video may be distracting. *Cruz* provides a complete solution to the problem of searching multimedia databases by teaching a user-centered interface for querying distributed multimedia databases. A user enters query keywords into an interface. The interface includes optional fields to allow the user to select a maximum number of objects to return, desired information sources, types of objects to display, and level of interaction.

*Cruz* also states:

On the initial screen (see Figure 2), the query keywords are specified and optional fields for customization are available. Keywords are entered as text, as in most engines, but unlike in most, the Boolean operators are provided. The operators are laid out to prevent their incorrect use and to eliminate the need for users to understand Boolean query construction.

The optional fields allow users to select the maximum number of objects to return, desired information sources, predefined page format (Section 2.3), type of objects to display, and level of interaction. Objects are of type text, image, audio, or video. Users that have saved searches may also select to return to their previous search results.

*Cruz*, section 2.1.

As shown above, *Cruz* teaches a virtual document display where query results are presented in a virtual document, including objects of various types. *Cruz* teaches:

The virtual document display is used to present users' query results in a single format that users can browse without leaving the Delaunay<sup>MM</sup> site.

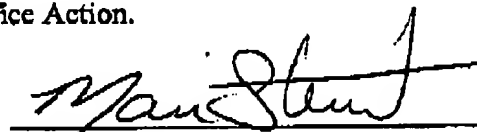
*Cruz*, section 2.3.

Thus, *Cruz* teaches presenting query results consisting of various types of media. However, *Cruz* does not teach or suggest extracting sets of text data associated in time with the plurality of moving video frames from the multimedia data and extracting a video frame from the first plurality of moving video frames associated with the first text data set to form a still image, as is claimed in independent claims 1, 8, and 15. In view of the above, *Loui*, *Bergen*, and *Cruz*, taken either alone or in combination, fail to teach or suggest the specific features recited in independent claims 1, 8, and 15, from which claims 7, 14, and 21 depend.

Moreover, *Cruz* does not teach or suggest discarding remaining moving video frames from the first plurality of moving video frames, as is recited in claims 7, 14, and 21. The Examiner alleges that *Cruz* teaches this feature in Figure 2, page 593, because *Cruz* teaches a "Video" checkbox. Applicants respectfully disagree. Deselecting the "Video" checkbox in Figure 2 of *Cruz* would not result in discarding remaining moving video frames after extracting a still image from the moving video frames. Rather, deselecting the "Video" checkbox would result in querying media sources that are not video at all. Therefore, the applied references fail to teach each and every claim limitation and, thus, fail to render claims 7, 14, and 21 obvious. Accordingly, Appellants respectfully request that the rejection of claims 7, 14, and 21 under 35 U.S.C. § 103(a) not be sustained.

### CONCLUSION

In view of the above, Appellant respectfully submits that claims 1, 3-8, 10-15 and 17-21 are allowable over the cited prior art and that the application is in condition for allowance. Accordingly, Appellant respectfully requests the Board of Patent Appeals and Interferences to not sustain the rejections set for the in the Final Office Action.



Mark Stewart  
Reg. No. 37,995  
YEE & ASSOCIATES, P.C.  
PO Box 802333  
Dallas, TX 75380  
(972) 385-8777



**CLAIMS APPENDIX**

The text of the claims involved in the appeal are:

1. A method for presenting text from moving video to a user, the method comprising:  
receiving multimedia data containing a plurality of moving video frames and an associated plurality of sets of text data, wherein the associated plurality of sets of text data are associated in time with the plurality of moving video frames, wherein the plurality of sets of text data includes a first text data set associated with a first plurality of moving video frames of the multimedia data, and a second text data set associated with a second plurality of moving video frames of the multimedia data;  
extracting the associated plurality of sets of text data from the multimedia data;  
extracting a first video frame, from the first plurality of moving video frames, associated with the first text data set to form a first still image;  
extracting a second video frame, from the second plurality of moving video frames, associated with the first text data set to form a second still image;  
outputting the first text data set in association with the first still image; and  
outputting the second text data set in association with the second still image.
3. The method as recited in claim 1, wherein the first text data set and the second text data set are presented in association with the first still image and the second still image, respectively, to the user simultaneously.

4. The method as recited in claim 3, wherein the first text data set and the second text data set are presented in association with the first still image and the second still image, respectively, in separate portions of a static display.
5. The method as recited in claim 1, wherein the first text data set and the second text data set are presented in association with the first still image and the second still image, respectively, to the user individually in a sequential order.
6. The method as recited in claim 5, wherein a next set of text data in the sequential order is presented in response to an indication by the user to display the next set of text data.
7. The method as recited in claim 1, wherein the step of extracting the associated plurality of sets of text data comprises parsing the multimedia data to determine the first text data set and the first video frame of the first plurality of moving video frames and discarding remaining moving video frames from the first plurality of moving video frames.
8. A computer program product in a computer readable media for use in a data processing system for presenting text from moving video to a user; the computer program product comprising:  
  
instructions for receiving multimedia data containing a plurality of moving video frames and an associated plurality of sets of text data, wherein the associated plurality of sets of text data are associated in time with the plurality of moving video frames, wherein the plurality of sets of text data includes a first text data set associated with a first plurality of moving video

frames of the multimedia data, and a second text data set associated with a second plurality of moving video frames of the multimedia data;

instructions for extracting the associated plurality of sets of text data from the multimedia data;

instructions for extracting a first video frame, from the first plurality of moving video frames, associated with the first text data set to form a first still image;

instructions for extracting a second video frame, from the second plurality of moving video frames, associated with the first text data set to form a second still image;

instructions for outputting the first text data set in association with the first still image;  
and

instructions for output the second text data set in association with the second still image.

10. The computer program product as recited in claim 8, wherein the first text data set and the second text data set are presented in association with the first still image and the second still image, respectively, to the user simultaneously.

11. The computer program product as recited in claim 10, wherein the the first text data set and the second text data set are presented in association with the first still image and the second still image, respectively, in separate portions of a static display.

12. The computer program product as recited in claim 8, wherein the first text data set and the second text data set are presented in association with the first still image and the second still image, respectively, to the user individually in a sequential order.

13. The computer program product as recited in claim 12, wherein a next set of text data in the sequential order is presented in response to an indication by the user to display the next set of text data.

14. The computer program product as recited in claim 8, wherein the instructions for extracting the associated plurality of sets of text data from the multimedia data comprise instructions for parsing the multimedia data to determine the first text data set and the first video frame of the first plurality of moving video frames and discarding remaining moving video frames from the first plurality of moving video frames.

15. A system for presenting text from moving video to a user; the system comprising:

a receiver which receives multimedia data containing a plurality of moving video frames and an associated plurality of sets of text data, wherein the associated plurality of sets of text data are associated in time with the plurality of moving video frames, wherein the plurality of sets of text data includes a first text data set associated with a first plurality of moving video frames of the multimedia data, and a second text data set associated with a second plurality of moving video frames of the multimedia data;

a text extraction unit which extracts the associated plurality of sets of text data from the multimedia data;

a still image extraction unit which extracts a first video frame, from the first plurality of moving video frames, associated with the first text data set to form a first still image and extracts a second video frame, from the second plurality of moving video frames, associated with the first text data set to form a second still image; and

an output unit which outputs the first text data set in association with the first still image and outputs the second text data set in association with the second still image.

17. The system as recited in claim 15, wherein the first text data set and the second text data set are presented in association with the first still image and the second still image, respectively, to the user simultaneously.

18. The system as recited in claim 17, wherein the first text data set and the second text data set are presented in association with the first still image and the second still image, respectively, in separate portions of a static display.

19. The system as recited in claim 15, wherein the first text data set and the second text data set are presented in association with the first still image and the second still image, respectively, to the user individually in a sequential order.

20. The system as recited in claim 19, wherein a next set of text data in the sequential order is presented in response to an indication by the user to display the next set of text data.

21. The system as recited in claim 15, wherein the extraction unit parses the multimedia data to determine the first text data set and the first video frame of the first plurality of moving video frames and discards remaining moving video frames from the first plurality of moving video frames.

**EVIDENCE APPENDIX**

There is no evidence to be presented.

**RELATED PROCEEDINGS APPENDIX**

There are no related proceedings.